ELSEVIER

Chemosphere 60 (2005) 898–906

CHEMOSPHERE

www.elsevier.com/locate/chemosphere

Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study

Andrea Baccarelli [a,b,*], Ruth Pfeiffer [a], Dario Consonni [b], Angela C. Pesatori [b], Matteo Bonzini [b], Donald G. Patterson Jr. [c], Pier Alberto Bertazzi [b], Maria Teresa Landi [a]

[a] *Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, 6120 Executive Boulevard, Bethesda, MD 20852, USA*

[b] *Department of Occupational and Environmental Health, EPOCA Research Center for Occupational, Clinical and Environmental Epidemiology, University of Milan, Via San Barnaba 8, 20122 Milan, Italy*

[c] *Division of Environmental Health Laboratory Science, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Highway, NE Atlanta, GA 30341, USA*

Received 1 October 2004; received in revised form 30 December 2004; accepted 14 January 2005
Available online 2 March 2005

Abstract

Exposure measurements of concentrations that are non-detectable or near the detection limit (DL) are common in environmental research. Proper statistical treatment of non-detects is critical to avoid bias and unnecessary loss of information. In the present work, we present an overview of possible statistical strategies for handling non-detectable values, including deletion, simple substitution, distributional methods, and distribution-based imputation. Simple substitution methods (e.g., substituting 0, DL/2, DL/$\sqrt{2}$, or DL for the non-detects) are the most commonly applied, even though the EPA Guidance for Data Quality Assessment discouraged their use when the percentage of non-detects is >15%. Distribution-based multiple imputation methods, also known as robust or "fill-in" procedures, may produce dependable results even when 50–70% of the observations are non-detects and can be performed using commonly available statistical software. Any statistical analysis can be conducted on the imputed datasets. Results properly reflect the presence of non-detectable values and produce valid statistical inference. We describe the use of distribution-based multiple imputation in a recent investigation conducted on subjects from the Seveso population exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), in which 55.6% of plasma TCDD measurements were non-detects. We suggest that distribution-based multiple imputation be the preferred method to analyze environmental data when substantial proportions of observations are non-detects.

* Corresponding author. Address: Department of Occupational and Environmental Health, EPOCA Research Center for Occupational, Clinical and Environmental Epidemiology, University of Milan, Via San Barnaba 8, 20122 Milan, Italy. Tel.: +39 02 503 20127; fax: +39 02 503 20126.

E-mail address: andrea.baccarelli@unimi.it (A. Baccarelli).

0045-6535/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2005.01.055

Keywords: 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin; Exposure assessment; Multiple imputation; Non-detects; Seveso; Detection limit

1. Introduction

Environmental research frequently relies on measurements of chemical, physical or biological agents performed to evaluate low-level contamination of air, soil, water or food, and to quantify the exposure of wildlife and human individuals. In spite of extensive efforts to develop high-sensitivity assays, often a substantial proportion of samples have such low concentrations to border on the detection limit (DL) defined by the sampling and analytical methods. Uncertainty deriving from levels that are non-detectable may impair the capability of drawing conclusions functional to regulatory decision making (Currie, 2000). Dioxins, which may pose a threat to human health and the environment even at very low concentrations, often challenge investigators with exposure measurements including high proportions of non-detects (Currie, 2000; Singh and Nocerino, 2002).

Also in recent environmental investigations, percentages of non-detectable levels in environmental and biological samples have often been large, as presence of more than 40% of non-detects for at least one of the analytes investigated has been far from being a rare occurrence (Acquavella et al., 2004; Barra et al., 2004; Berkowitz et al., 2004; Caserini et al., 2004; Kato et al., 2004; Liu and Mou, 2004; Quandt et al., 2004; Roots et al., 2004; Silva et al., 2004; Sinkkonen et al., 2004; Toro et al., 2004). However, in spite of intense debate and extensive theoretical research activity on the topic, environmental research has often tolerated the loss of information and potential bias arising from improper or inadequate treatment of non-detects and rarely taken advantage of available statistical techniques to limit these problems.

In the present work, we discuss possible strategies for handling exposure data including non-detects. We recommend the use of a multiple imputation method based on distribution-based estimation of non-detectable values. Performances of this method have been previously assessed through data simulation studies (Helsel, 1990; Huybrechts et al., 2002; Lubin et al., 2004). We show its application in estimating mean levels of dioxin in subjects sampled from the Seveso population exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD). In this recent study, 55.6% of the subjects had non-detectable plasma TCDD levels.

2. Non-detects: what they are and how to handle them

Measurements including non-detects are ''left-censored data''. A left-censored data sampling distribution is one for which the only information in some of the samples is that the true measurement is less than a given censoring value (Hornung and Reed, 1990). The censoring value, i.e., the DL, may be constant for all the observations in the dataset (singly censored data) or may vary between observations (multiply censored data). There have been considerable differences of opinion about how to define the detection limit and how to determine it experimentally (Currie, 2000). For our purposes, we will simply define DL as the minimum level of an analyte in a given determination that can reliably be reported as an accurate number (Adams, 1998). Although non-detects do not provide a point measure of the analyte, they are informative data indicating that the analyte has concentration between 0 and DL.

There are a number of available techniques for treating non-detects (Table 1). We enumerate and discuss those most commonly used, with particular reference to their possible application in environmental research.

2.1. Deletion

The simplest approach is to discard all the observations with non-detectable levels. This procedure causes loss of the information carried by non-detects and excludes from the dataset the samples with lowest concentrations. Consequently, the mean analyte levels, which are based on the remaining above-DL observations, are overestimated (Hornung and Reed, 1990).

2.2. Simple substitution

Simple substitution methods substitute a single value chosen from the interval zero to DL (e.g., zero, DL/2, DL/$\sqrt{2}$, or DL) for each of the non-detects. Summary statistics are then calculated using these substitute numbers together with the values above DL. Although widely used, it has been noted that these methods have no theoretical basis and perform poorly compared to other procedures (Helsel, 1990). The substitution of zero produces means that are underestimates of the true means, whereas substitution of the DL value causes mean estimates to be biased upward. Hornung and Reed (1990) found that, for any censoring proportion and degree of variability in the simulated data they examined, the DL/2 or DL/$\sqrt{2}$ substitutions tended to be less precise than other more complex procedures, such as the distributional methods described in the next section. Although DL/2 and DL/$\sqrt{2}$ substitutions are sometimes indicated as not requiring any assumption on the underlying distribution of the data, it has been shown that they are actually based on the implicit hypothesis that values

Table 1
Available methods for the treatment of measurements data in the presence of non-detectable values

Procedure	Method	Validity		Uncertainty due to non-detectable values
		Mean	Standard deviation	
Deletion	Non-detects discarded	Overestimated	Underestimated	Unaccounted for
Simple substitution				
Zero	Non-detects set to zero	Underestimated	Overestimated	Unaccounted for[a]
DL/2	Non-detects set to half the detection limit	Bias small if • frequency of non-detects is low • highly skewed data		Unaccounted for[a]
DL/$\sqrt{2}$	Non-detects set to the detection limit divided by $\sqrt{2}$	Bias small if • frequency of non-detects is low • not highly skewed data		Unaccounted for[a]
DL	Non-detects considered equal to the detection limit	Overestimated	Underestimated	Unaccounted for[a]
Distributional	Mean and SD estimated using assumptions on underlying data distribution	Bias small if • actual data do not depart from the assumed distribution • <50–60% of non-detects		Unaccounted for
Distribution-based imputation	Imputes a value drawn from assumed underlying distribution	Unbiased even if • data show mild/moderate departure from the assumed distribution • 60–70% of non-detects		Accounted for by multiple imputation

Abbreviation: DL, detection limit.
[a] Substituted values are treated in the analysis as true measured concentrations.

below-DL follow a uniform (DL/2) or triangular distribution (DL/$\sqrt{2}$, approximating the left tail of a lognormal distribution) (Hornung and Reed, 1990). The EPA Guidance for Data Quality Assessment suggests that simple substitution methods may be adequate only when the percentage of non-detects is low (<15%), and should be avoided for higher percentages (EPA, 2000).

2.3. Distributional methods

Distributional methods are based on the assumption that data arise from a specified parametric distribution, e.g., the lognormal. Parameters of the distribution (e.g., mean and standard deviation) are estimated for example, by maximum-likelihood estimation (MLE), based on observed concentrations and DL values (Huybrechts et al., 2002). Distributional methods perform well only when the true distribution of the observations corresponds to the assumed distribution and the proportion of non-detects is <50–60% (Table 1). Although, in the past, distributional methods were "laborious, requiring extensive calculations and the use of tables" (Hornung and Reed, 1990), desktop computing has greatly decreased those difficulties and standard software is readily available (Finkelstein and Verma, 2001).

2.4. Distribution-based imputation

2.4.1. Imputation for non-detectable values

Distribution-based imputation procedures, referred to by some authors as "robust" or "fill-in" methods (Helsel, 1990), fit a parametric distribution to the data using the same statistical procedures employed by distributional methods. Then the fitted distribution is used to "draw" a value for each of the non-detects so that a complete dataset is created that can be used in the analysis. While distributional methods are unbiased only when the true distribution of the observations corresponds to the assumed distribution, imputation procedures are robust to mild or moderate departures of the data from the assumed distributional shape (Huybrechts et al., 2002). In addition, imputation procedures may generate accurate estimates of population parameters even when the percentage of non-detects is as high as 60–70% (Table 1) (Huybrechts et al., 2002; Lubin et al., 2004).

2.4.2. Multiple imputations for statistical inference

When using simple substitution or distributional methods, statistical estimates are usually treated as if they were calculated on actual measured data and uncertainty resulting from substitution or imputation is mostly ignored (Table 1). By contrast, the multiple imputation strategy replaces each non-detect with several values that represent the uncertainty about which value to impute (Rubin, 1976, 1987). Multiple datasets are created by repeating the imputation for non-detects and, from each of them, parameter estimates and covariances are obtained using standard analysis. These estimates are then combined and the total variance of the final estimate is computed (Rubin, 1987). When the underlying distribution of the data is known, this strategy results in valid statistical inference that properly reflects the uncertainty due to imputed values.

2.5. Remarks

All methods reported in Table 1 may be used with both singly or multiply censored data (Helsel and Cohn, 1988; Helsel, 1990). Handling of measurements including non-detectable values in developing regression models to examine the relationship between the measurement value and covariate factors is examined in detail by Lubin et al. (2004). We just mention here that distribution-based imputation procedures also allow for the use of regression methods on datasets completed by the imputed data, e.g., to evaluate a dose–response relationship between the exposure and the outcome of interest, or to adjust for possible confounders.

3. Application of distribution-based multiple imputation and comparison with other methods in the Seveso chloracne study

3.1. Study background

In 1976, the Seveso accident exposed a large residential population to TCDD, the most toxic dioxin congener. The exposure produced a large outbreak of chloracne, mostly among children (Baccarelli et al., in press). Chloracne is a skin intoxication similar in appearance to *acne vulgaris*, but characterized by pale-yellow keratin cysts and larger and prominent comedones. After the accident, the area was divided in four zones of decreasing contamination: zone A, where TCDD soil concentration was highest, with 723 inhabitants; zone B with 4281 inhabitants; zone R, with 31 643 inhabitants; and a non-contaminated area surrounding the contaminated zones, which had 181 576 inhabitants. Between 1993 and 1998, we contacted 101 chloracne cases (56 males, 45 females; median age at the accident 8 years, range: 6 months–46 years) with confirmed diagnoses and 211 controls (108 males, 103 females; median age at the accident 14 years, range: 3 months–58 years) selected from the Seveso population (Baccarelli et al., in press). TCDD levels were measured at the Centers for Disease Control and Prevention (CDC) using a high-resolution gas chromatography/high resolution mass spectrometric analysis performed on human plasma. All measurements were performed in the same testing facility by using the same assay, technology, and standardized procedures, as previously described (Patterson et al., 1986, 1987). Briefly, criteria for a positive TCDD determination were as follows: (1) signal/noise greater than 3/1 for both signals on ions 320 and 322; (2) signal/noise greater than 10/1 for both signals on ions 332 and 334 from the internal standard; (3) observed retention times within ±1 scan of each other on ions 320 and 322 and the relative retention time (RRT) (to [13C12]-2,3,7,8-TCDD) within 2 part-per-thousand of the RRT of the analytical standard; (4) ratios of the intensities of the ion 320–322 and 332–334 within the 95% confidence intervals established for these ratios (Patterson et al., 1987). Details on criteria for reporting results as quantified, non-quantifiable, or non-detectable, including how responses falling outside of the theoretical ion-abundance ratio were handled, were previously described (Patterson et al., 1986). Nineteen of the 312 subjects had samples inadequate for the assay and were excluded from the analysis. On average, 5 ml of plasma were used for the assay. The assay has variable DL that depends, among other factors, on the amount of lipids that can be extracted from the samples. The amount of lipid available is related to the amount of serum available for the assay. In addition, the recovery of the analyte through the cleanup procedure from the lipid can vary from sample to sample and this will also affect the DL for that sample. Plasma TCDD levels were non-detectable in 163 (55.6%) of the subjects. Subjects with non-detectable levels had a median DL equal to 6.5 pg g^{-1} fat (range 1.4–26.8).

3.2. Distribution-based imputation of non-detects

TCDD plasma levels, similar to most environmental pollutants (EPA, 2000; Huybrechts et al., 2002), are well approximated by a lognormal distribution (Papke et al., 1996; Landi and Baccarelli, 2003). First, we set an upper bound at the DL reported for each sample. To obtain maximum likelihood estimates of the mean and variance of the lognormal distribution, we used the same likelihood in formula (1) of the article by Huybrechts et al. (2002). This likelihood function was based on the log-transformed TCDD data from the control subjects, and was the product of the normal density function for values of the above-DL measurements, and the cumulative normal distribution function at the log-transformed bounds for the non-detects. We only used

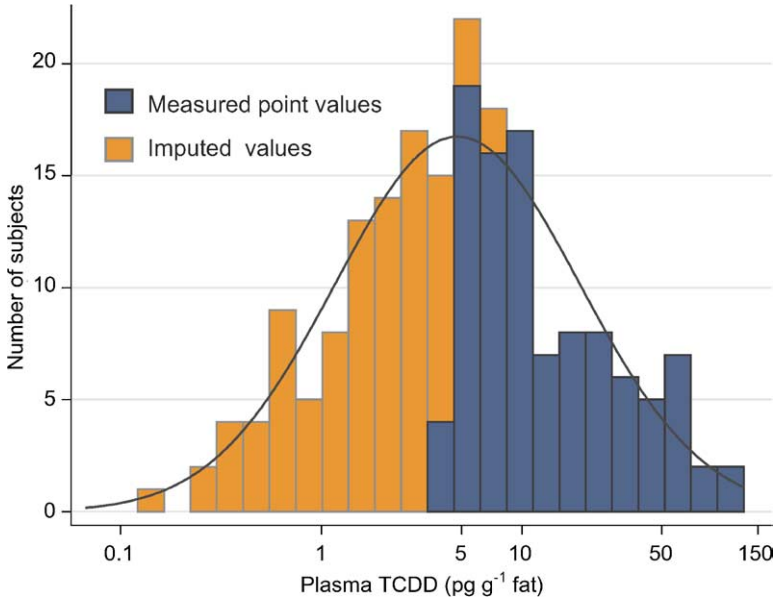

Fig. 1. Measured data point and imputed values of plasma 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) (pg g^{-1} fat or ppt, lipid adjusted) among controls in one of the 10 imputations of non-detectable TCDD values in the Seveso chloracne study.

control subjects to estimate the parameters of the lognormal distribution, because they are more representative of the general population, while chloracne cases were relatively rare in the Seveso population and are likely to have higher dioxin levels (Baccarelli et al., in press; Needham et al., 1997). We then imputed a value from the lognormal distribution with the estimated mean and variance parameters for each non-detect, by drawing a random observation from this distribution conditional on the observation being lower than DL. Fig. 1 shows an example of a complete dataset including actual above-DL measured values and imputed values for subjects with non-detectable TCDD levels.

The above process produced a complete dataset with the desirable quality that the estimates of population parameters obtained from it are unbiased assuming the correct distribution was chosen. To incorporate uncertainty resulting from the imputation, we followed the approach by Rubin (1976, 1987) and repeated the imputation and estimation 10 times. For each of the 10 datasets completed by imputing for non-detectable values, we estimated associations of TCDD plasma levels with case status or other variables using linear regression models. We assessed the association between chloracne and TCDD plasma levels by fitting logistic regression models. All analysis were adjusted for gender, and age and residence at the time of the accident. The resulting estimates were combined using PROC MIANALYZE in SAS 8.2 (SAS Institute Inc., 2001). The combined final estimate obtained from multiple imputations is the average of the 10 complete-data estimates. The total variance of the estimate is the sum of the within-imputation variance and the between-imputation variance. The within-imputation variance is defined as the average of the complete-data variances, and the between-imputation variance is the sample variance of the 10 complete dataset estimates (Rubin, 1987). Unless proportions of values that need to be imputed are extremely high, there is little or no practical benefit from using more than 10 imputations (Schafer, 1999).

To obtain means by case status, gender, age groups, or body mass index (BMI) categories, we combined the coefficients obtained from linear regression models to estimate stratum-specific TCDD levels, assuming that the value of the other covariates in the models was constant and equal to their means. As the plasma TCDD distribution is lognormal, we calculated geometric, rather than arithmetic, means.

3.3. Mean plasma TCDD in chloracne cases and controls, by using distribution-based multiple imputation

We first used the distribution-based imputation procedure to estimate the mean levels of plasma TCDD in cases and controls (Table 2). In the 10 datasets obtained by multiple imputation for non-detects, geometric means of plasma TCDD, adjusted for age, gender and zone of residence, varied between 7.3 and 9.4 pg g^{-1} fat in chloracne cases and 4.1–5.1 pg g^{-1} fat in controls. The variation across the 10 datasets reflects the uncertainty due to the estimation of the non-detectable values. In all of the datasets the difference between cases and

Table 2
Plasma TCDD levels estimated using distribution-based multiple imputation for non-detectable values in chloracne subjects and controls from the Seveso population

Results	Plasma TCDD (pg g^{-1} fat)[a]					Relative risk of chloracne	
	Chloracne cases (n = 98)		Control subjects (n = 195)		p-Value	OR[b,c]	(95% CI)[c]
	Geometric mean[c]	(Min–Max)	Geometric mean[c]	(Min–Max)			
Final[d]	8.6	(0.3–475.0)	4.6	(0.3–127.0)	0.002	1.37	(1.17–1.59)
Imputation 1	9.2	(0.1–475.0)	4.5	(0.1–127.0)	<0.001	1.38	(1.18–1.61)
Imputation 2	9.4	(0.4–475.0)	5.1	(0.3–127.0)	<0.001	1.37	(1.17–1.59)
Imputation 3	7.8	(0.1–475.0)	4.5	(0.1–127.0)	0.003	1.36	(1.17–1.58)
Imputation 4	8.4	(0.1–475.0)	4.7	(0.1–127.0)	0.001	1.37	(1.17–1.59)
Imputation 5	8.6	(0.2–475.0)	5.0	(0.1–127.0)	0.001	1.35	(1.16–1.57)
Imputation 6	8.7	(0.2–475.0)	4.1	(0.1–127.0)	<0.001	1.38	(1.18–1.61)
Imputation 7	8.1	(0.1–475.0)	4.9	(0.2–127.0)	0.004	1.36	(1.17–1.59)
Imputation 8	9.1	(0.4–475.0)	4.3	(0.2–127.0)	<0.001	1.37	(1.17–1.60)
Imputation 9	7.3	(0.3–475.0)	4.4	(0.3–127.0)	0.007	1.35	(1.16–1.57)
Imputation 10	9.1	(0.1–475.0)	5.0	(0.2–127.0)	<0.001	1.38	(1.18–1.62)

Abbreviations: TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; OR, Odds ratio; CI, confidence interval.

[a] Lipid-adjusted plasma TCDD levels measured approximately 20 years after the accident. Non-detects were found in 55.6% of the subjects.

[b] Relative odds of chloracne for each 10 pg g^{-1} fat increase in plasma TCDD.

[c] Adjusted for gender, and age and zone of residence at the time of the accident.

[d] Obtained from the results of the 10 imputations according to Rubin (1987) using PROC MIANALYZE of SAS 8.2 (see Section 2.4.1).

controls was statistically significant ($p \leqslant 0.007$). In the final results, obtained by combination of the analyses conducted on the 10 imputed datasets (Rubin, 1987), we estimated mean plasma TCDD levels equal to 8.6 pg g^{-1} fat in cases and 4.6 pg g^{-1} fat in controls ($p = 0.002$ for the difference between cases and controls). Note that the final p-value is larger than most p-values obtained from individual imputations, as it incorporates the between-imputation variability, reflecting the uncertainty related to the lack of point-measures for the non-detects.

We also evaluated the dose–response relationship between chloracne occurrence and plasma TCDD levels (Table 2). The final estimates showed that each increase of 10 pg g^{-1} fat in plasma TCDD was associated with a 37% elevation of the relative odds of chloracne (OR = 1.37, 95% CI: 1.17–1.59). The ORs for chloracne showed small variations across different imputations (lowest OR, = 1.35, 95% CI 1.16–1.57; highest OR = 1.38, 1.18–1.62).

3.4. Comparison with other methods

We repeated the estimation of mean plasma TCDD in chloracne cases and controls using the other procedures reported in Table 1 and calculated the percentage of departure (relative bias) from the estimates obtained using the distribution-based multiple imputation method (reference) (Table 3).

Zero substitution and distributional methods could not be used in our study. The zero substitution does not allow to log-transform the data, a step necessary to compute geometric means. The distributional methods do not allow for adjustment by age, gender and zone of residence, as they do not accommodate regression models for mean estimation (Lubin et al., 2004).

As expected, excluding the non-detects (deletion) yielded geometric means that were much higher than the reference (Table 3). The relative bias was 329.6% in cases and 197.9% in controls. Means obtained using the remaining simple substitution methods were also higher than the reference. The relative bias varied between 22.8% and 82.7%, with relatively better results when the DL/2 substitution was used. Using deletion or substitution methods, the p-value for difference between cases and controls was always statistically significant ($p < 0.001$). The highly statistically significant p-values reflect, at least in part, the underestimation of total variance that occurs when these methods are used.

3.5. Determinants of TCDD plasma levels

We also used the distribution-based multiple imputation method to assess the association between TCDD and possible determinants of dioxin levels, such as proximity to the site of the accident, older age, female gender and BMI (Landi et al., 1997) (Table 4). We found that mean plasma TCDD increased from 2.9 pg g^{-1} fat in

Table 3
Comparison of different statistical procedures for treating non-detects in the chloracne case-control study

Procedure	Plasma TCDD (pg g^{-1} fat)[a]			
	Chloracne cases ($n = 98$)		Control subjects ($n = 195$)	
	Geometric mean[a] (pg g^{-1} fat)	Percent difference (%)[b]	Geometric mean[a] (pg g^{-1} fat)	Percent difference[b] (%)
Distribution-based multiple imputation	8.6	Reference	4.6	Reference
Deletion	36.9	329.6	13.8	197.9
Simple substitution				
Zero[c]	NA	–	NA	–
DL/2	10.6	23.9	5.7	22.8
DL/$\sqrt{2}$	12.8	49.2	6.9	49.8
DL	15.4	79.6	8.5	82.7
Distributional[d]	NA	–	NA	–

Abbreviations: DL, detection limit; TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; NA, not applicable.
[a] Adjusted for gender, age and zone of residence using multiple regression analysis.
[b] Percentage difference relative to the means calculated using the distribution-based multiple imputation procedure (reference).
[c] Log-transformation of values set to zero is impossible. Thus, geometric means could not be calculated.
[d] Distributional methods do not allow for calculation of geometric means adjusted through regression methods.

Table 4
Plasma TCDD levels by zone of residence, age, gender and body mass index, estimated using multiple imputation for non-detectable values

	n	Plasma TCDD[a]		
		Geometric mean	Min–Max	*p*-Value[a]
Residence at the time of the accident				
Non-contaminated area	77	2.9	(0.4–18.1)	Reference
R zone	76	3.7	(0.3–23.5)	0.38
B zone	63	5.8	(0.3–51.5)	<0.01
A zone	77	16.6	(0.3–475.0)	<0.001
Age at the time of the accident				
⩽8 years[b]	105	3.9	(0.3–122.0)	Reference
>8 years[b]	188	7.0	(0.3–475.0)	<0.001
Plasma TCDD percentage increase for each 10-year increment of age		+32.4%		<0.001
Gender				
Male	153	4.2	(0.3–104.0)	Reference
Female	140	7.9	(0.3–475.0)	<0.001
Body mass index (kg m^{-2})[c]				
<21.6	97	4.4	(0.3–301.0)	Reference
21.7–25	95	6.0	(0.3–447.0)	0.12
>25	99	7.0	(0.5–475.0)	0.04
Plasma TCDD percentage increase for each 1-kg m^{-2} increment of BMI		+5.5%		0.01

Abbreviations: TCDD, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin; BMI, body mass index.
[a] Geometric means (pg g^{-1} fat) and *p*-values computed using multiple linear regression models that included gender, and age and zone of residence at the time of the accident.
[b] Median age of the chloracne cases.
[c] Body mass index measured at the interview.

subjects from the non-contaminated area, to 3.7 pg g^{-1} fat in zone R, the zone with the lowest contamination ($p = 0.38$ vs. non-contaminated area); 5.8 pg g^{-1} fat in zone B ($p < 0.01$ vs. non-contaminated area); and

16.6 pg g^{-1} fat in zone A, the most contaminated zone ($p < 0.001$ vs. non-contaminated area; $p < 0.001$ for trend across zones). Individuals who were 8 years or younger at the accident had mean plasma TCDD equal to 3.9 pg g^{-1} fat (range 0.3–122.0), while TCDD plasma levels were higher in older subjects (mean = 7.0 pg g^{-1} fat, range 0.3–475.0) ($p < 0.001$). Each 10-year age difference was associated with a 32.4% increase in plasma TCDD ($p < 0.001$ for linear trend by age). Mean plasma TCDD levels were 4.2 pg g^{-1} fat (range 0.3–104.0) in males and 7.9 pg g^{-1} fat in females (range 0.3–475.0) ($p < 0.001$). Subjects with lower BMI (<21.6 kg m^{-2}) had mean plasma TCDD equal to 4.4 pg g^{-1} fat (range 0.3–301.0), which tended to increase in individuals with BMI between 21.7 and 25 kg m^{-2} (mean = 6.0 pg g^{-1} fat, range 0.3–447.0) and was highest in those above 25 kg m^{-2} (mean = 7.0 pg g^{-1} fat, 0.5–475.0). Each 1 kg m^{-2} increment of BMI contributed a 5.5% increase to plasma TCDD ($p = 0.01$ for linear trend by BMI).

4. Conclusions

Correct handling of non-detectable values is critical in environmental research, particularly when the range of analytes in the study is close to DL, as often occurs for persistent organic pollutants and dioxins. In our application on chloracne, the multiple-imputation means were considered as the best obtainable estimates (Huybrechts et al., 2002) and set as reference in the comparison with the other procedures, whose relative bias varied between 22.8% and 329.6%. Independently of the procedures we used, we observed a statistically significant difference between chloracne cases and controls in plasma TCDD. This reflects the clear difference in plasma TCDD between cases and controls at the higher end of the TCDD distribution, where some subjects, nearly all cases, had concentrations two orders of magnitude greater than the average DL. However, *p*-values for the case-control comparison obtained using deletion and simple substitution methods were lower than the *p*-value produced from multiple imputation, reflecting the underestimation of total variance that occurs when the simpler methods are used. Using multiple imputation methods, investigators obtain valid inference with lower probability of false positive results.

In dealing with lognormally distributed data, geometric means are usually reported, because they assign similar weight to all observations. Geometric means thus tend to be sensitive to how non-detects are handled. Arithmetic means, by contrast, are not strongly influenced by low-exposure levels, but are primarily sensitive to the larger values of exposure.

An additional advantage of imputation methods is that, once distribution-based values are imputed for the non-detectable values, any statistical method appropriate for complete data can be used. In the Seveso chloracne example, we showed the use of multiple linear regression to calculate adjusted geometric means and of multiple logistic regression models to evaluate the dose–response relationship of plasma TCDD with chloracne and determinants of TCDD body burden. As pointed out in our application, simpler methods may preclude investigators from the use of standard statistical techniques. Zero substitution and distributional methods, for example, could not be used for estimation of adjusted TCDD means in the Seveso data.

We emphasize that, in planning a study and performing laboratory measurements, maximum effort should be given to reduce the occurrence of non-detects. The sensitivity of the assay and sample volume required should be carefully evaluated in relation to the range of contamination investigated. When measurement data include non-detects, distribution-based imputation strategies may have definite advantages over other methods in handling measurements of persistent organic pollutants and dioxins, as well as other environmental contaminants. The computational effort required by the procedure has been reduced by the availability of standard software and is largely offset by the advantages that can be attained.

Acknowledgment

We are indebted to Jay Lubin, Ph.D. for the original inspiration of the present work, advice, and critical reading of the manuscript.

References

Acquavella, J.F., Alexander, B.H., Mandel, J.S., Gustin, C., Baker, B., Chapman, P., et al., 2004. Glyphosate biomonitoring for farmers and their families: results from the Farm Family Exposure Study. Environ. Health Perspect. 112, 321–326.

Adams, N.H., 1998. Quality assurance reviews: how they differ from peer reviews. Qual. Assur. 6, 75–85.

Baccarelli, A., Pesatori, A.C., Consonni, D., Mocarelli, P., Patterson, D., Caporaso, N.E., et al., in press. Health status and plasma dioxin levels in chloracne cases twenty years after the Seveso, Italy accident. Brit. J. Dermatol.

Barra, R., Cisternas, M., Suarez, C., Araneda, A., Pinones, O., Popp, P., 2004. PCBs and HCHs in a salt-marsh sediment record from South-Central Chile: use of tsunami signatures and 137Cs fallout as temporal markers. Chemosphere 55, 965–972.

Berkowitz, G.S., Wetmur, J.G., Birman-Deych, E., Obel, J., Lapinski, R.H., Godbold, J.H., et al., 2004. In utero pesticide exposure, maternal paraoxonase activity, and head circumference. Environ. Health Perspect. 112, 388–391.

Caserini, S., Cernuschi, S., Giugliano, M., Grosso, M., Lonati, G., Mattaini, P., 2004. Air and soil dioxin levels at three

sites in Italy in proximity to MSW incineration plants. Chemosphere 54, 1279–1287.

Currie, L.A., 2000. Detection and quantification capabilities and the evaluation of low-level data: some international perspectives and continuing challenges. J. Radioanal. Nucl. Chem. 245, 145–156.

EPA, U.S., 2000. Guidance for data quality assessment. Practical methods for data analysis. Office of Environmental Information, EPA QA/G-9, QA00 Version Washington, DC.

Finkelstein, M.M., Verma, D.K., 2001. Exposure estimation in the presence of nondetectable values: another look. AIHAJ 62, 195–198.

Helsel, D.R., 1990. Less than obvious-statistical treatment of data below the detection limit. Environ. Sci. Technol. 24, 1766–1774.

Helsel, D.R., Cohn, T.A., 1988. Estimation of descriptive statistics for multiply censored water-quality data. Water Resour. Res. 24, 1997–2004.

Hornung, R.W., Reed, L.D., 1990. Estimation of average concentration in the presence of non-detectable values. Appl. Occup. Environ. Hyg. 5, 48–51.

Huybrechts, T., Thas, O., Dewulf, J., Van Langenhov, H., 2002. How to estimate moments and quantiles of environmental data sets with non-detected observations? A case study on volatile organic compounds in marine water samples. J. Chromatogr. A 975, 123–133.

Kato, K., Silva, M.J., Reidy, J.A., Hurtz III, D., Malek, N.A., Needham, L.L., et al., 2004. Mono(2-ethyl-5-hydroxyhexyl) phthalate and mono-(2-ethyl-5-oxohexyl) phthalate as biomarkers for human exposure assessment to di-(2-ethylhexyl) phthalate. Environ. Health Perspect. 112, 327–330.

Landi, M.T., Baccarelli, A., 2003. Correspondence re: Toide et al., Aryl Hydrocarbon Hydroxylase represents CYP1B1, and not CYP1A1, in human freshly isolated white cells: trimodal distribution of Japanese population according to induction of CYP1B1 mRNA by environmental dioxins. Cancer Epidemiol. Biomarkers Prev. 12, 1116–1117.

Landi, M.T., Needham, L.L., Lucier, G., Mocarelli, P., Bertazzi, P.A., Caporaso, N., 1997. Concentrations of dioxin 20 years after Seveso. Lancet 349, 1811.

Liu, Y., Mou, S., 2004. Determination of bromate and chlorinated haloacetic acids in bottled drinking water with chromatographic methods. Chemosphere 55, 1253–1258.

Lubin, J.H., Colt, J.S., Camann, D., Davis, S., Cerhan, J.R., Severson, R.K., et al., 2004. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ. Health Perspect. 112, 1691–1696.

Needham, L.L., Gerthoux, P.M., Patterson Jr., D.G., Brambilla, P., Turner, W.E., Beretta, C., et al., 1997. Serum dioxin levels in Seveso, Italy, population in 1976. Teratog. Carcinog. Mutagen. 17, 225–240.

Papke, O., Ball, M., Lis, A., Wuthe, J., 1996. PCDD/PCDFs in humans, follow-up of background data for Germany, 1994. Chemosphere 32, 575–582.

Patterson, D.G., Holler, J.S., Lapeza Jr., C.R., Alexander, L.R., Groce, D.F., O'Connor, R.C., et al., 1986. High-resolution gas chromatographic/high-resolution mass spectrometric analysis of human adipose tissue for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. Anal. Chem. 58, 705–713.

Patterson, D.G., Hampton, L., Lapeza Jr., C.R., Belser, W.T., Green, V., Alexander, L., et al., 1987. High-resolution gas chromatographic/high-resolution mass spectrometric analysis of human serum on a whole-weight and lipid basis for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. Anal. Chem. 59, 2000–2005.

Quandt, S.A., Arcury, T.A., Rao, P., Snively, B.M., Camann, D.E., Doran, A.M., et al., 2004. Agricultural and residential pesticides in wipe samples from farmworker family residences in North Carolina and Virginia. Environ. Health Perspect. 112, 382–387.

Roots, O., Henkelmann, B., Schramm, K.W., 2004. Concentrations of polychlorinated dibenzo-*p*-dioxins and polychlorinated dibenzofurans in soil in the vicinity of a landfill. Chemosphere 57, 337–342.

Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592.

Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley, New York.

SAS Institute Inc., 2001. The MIANALYZE procedure. In: SAS/STAT software: changes and enhancements. Release 8.2, SAS Institute Inc., Cary, NC, pp. 201–233.

Schafer, J.L., 1999. Multiple imputation: a primer. Stat. Methods Med. Res. 8, 3–15.

Silva, M.J., Barr, D.B., Reidy, J.A., Malek, N.A., Hodge, C.C., Caudill, S.P., et al., 2004. Urinary levels of seven phthalate metabolites in the U.S. population from the National Health and Nutrition Examination Survey (NHANES) 1999–2000. Environ. Health Perspect. 112, 331–338.

Singh, A., Nocerino, J., 2002. Robust estimation of mean and variance using environmental data sets with below detection limit observations. Chemometr. Intell. Lab. Syst. 60, 69–86.

Sinkkonen, S., Rantalainen, A.L., Paasivirta, J., Lahtipera, M., 2004. Polybrominated methoxy diphenyl ethers (MeO-PBDEs) in fish and guillemot of Baltic, Atlantic and Arctic environments. Chemosphere 56, 767–775.

Toro, B., Palma-Fleming, H., Navarro, J.M., 2004. Organic pollutant burden of the giant mussels Choromytilus chorus from the South-Central Chilean coast. Chemosphere 55, 267–275.